

# Application of Information Theory, Lecture 5

## Channel Capacity and Isoperimetric Inequality

### Handout Mode

Iftach Haitner

Tel Aviv University.

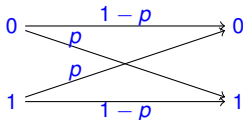
April 12, 2018

# Part I

## Channel Capacity

# The problem

- ▶ We want to send a message  $\mathbf{x} = (x_1, \dots, x_m) \in \{0, 1\}^m$ , but the communication channel is **faulty**
- ▶ Each bit is (independently) flipped w.p.  $p$  (e.g., 0.1)



- ▶ (expected) **Error rate** is  $p$
- ▶ Such “channel” is called **Binary Symmetric Channel (BSC)**
- ▶ When sending  $m$  bits, we have  $\approx pm$  errors
- ▶ Can we send bits with smaller error?

## Solution

- ▶ Obvious solution is “error correction codes (ECC)”
- ▶ Simplest example: send each bit three times, and take majority
- ▶ Error happens if the channel errs at least twice
- ▶ For  $p = 0.1$ : happens w.p.  
 $3p^2(1 - p) + p^3 = 3 \cdot 0.01 \cdot 0.9 + 0.001 = 0.028$
- ▶ Error rate: .028
- ▶ **Transmission rate**:  $1/3$  (i.e., # of bits recovered / #of bits transmitted)
- ▶ We reduced the error rate by reducing the transmission rate.
- ▶ Can we reduce the error rate, **without** reducing the transmitting rate too much?
- ▶ Before Shannon it was believed that very small error rate requires very small transmission rate.

# Shannon's result

- ▶ Shannon showed that you can reduce the error rate towards 0, without reducing the transmission rate towards 0
- ▶ For any  $c < C_p$ , exists a code with transmission rate  $c$  that is correct w.h.p.
- ▶ Example: for  $p = .1$ ,  $C_p > \frac{1}{2}$ .

Hence, for sending  $\mathbf{x} = (x_1, \dots, x_m)$ , one can send  $2m$  bits, such that  $\mathbf{x}$  is recovered w.p. close to 1

- ▶ More generally,  $\forall p \in [0, 1] \exists C_p$  such that for sending  $\mathbf{x} \in \{0, 1\}^m$ , one can send  $\approx \frac{m}{C_p}$  bits, and  $\mathbf{x}$  is recovered w.p. close to 1
- ▶  $C_p$  might be 0 (i.e., for  $p = \frac{1}{2}$ )
- ▶ A revolution in EE and the whole world

## Error correction code

- ▶ Message to send  $\mathbf{x} = (x_1, \dots, x_m) \in \{0, 1\}^m$
- ▶ Encoding scheme:  $f: \{0, 1\}^m \mapsto \{0, 1\}^n$  ( $n > m$ )
- ▶ Decoding scheme:  $g: \{0, 1\}^n \mapsto \{0, 1\}^m$
- ▶  $\frac{m}{n}$  — transmission rate
- ▶ Sender sends  $f(\mathbf{x})$  rather than  $\mathbf{x}$
- ▶ Receiver decodes the message by applying  $g$

$$\underbrace{\mathbf{x}}_{m \text{ bits}} \xrightarrow{\text{encoding}} \underbrace{f(\mathbf{x})}_{n \text{ bits}} \xrightarrow{\text{channel}} \underbrace{f(\mathbf{x}) \oplus Z}_{\text{bitwise XOR}} \xrightarrow{\text{decoding}} g(f(\mathbf{x}) \oplus Z)$$

$Z = (Z_1, \dots, Z_n)$  where  $Z_1, \dots, Z_n$  iid  $\sim (1 - p, p)$  (i.e., over  $\{0, 1\}$  with  $\Pr[Z_i = 1] = p$ )

- ▶ We hope  $g(f(\mathbf{x}) \oplus Z) = \mathbf{x}$
- ▶ ECCs are everywhere
- ▶ ECC Vs compression

# Shannon's theorem

## Theorem 1

$\forall p \quad \exists C_p, \text{ s.t. } \forall \varepsilon > 0 \quad \exists m_\varepsilon, \text{ s.t. } \forall m \geq m_\varepsilon \text{ and } n \geq m(\frac{1}{C_p} + \varepsilon),$   
 $\exists f: \{0, 1\}^m \mapsto \{0, 1\}^n \text{ and } g: \{0, 1\}^n \mapsto \{0, 1\}^m, \text{ s.t. } \forall \mathbf{x} \in \{0, 1\}^m:$   
$$\Pr_{z \leftarrow Z = (Z_1, \dots, Z_n)} [g(f(\mathbf{x}) \oplus z) \neq \mathbf{x}] \leq \varepsilon$$

for  $Z_1, \dots, Z_n \text{ iid} \sim (1 - p, p)$ .

- ▶  $C_p = 1 - h(p)$  — the **channel capacity**

$$p = .1 \implies C_p = 0.5310 > \frac{1}{2}$$

$$p = .25 \implies C_p \approx \frac{1}{5}$$

- ▶ Tight theorem
- ▶ We prove a weaker variant that holds w.h.p. over  $\mathbf{x} \leftarrow \{0, 1\}^m$

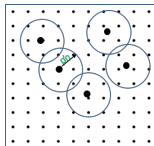
# Hamming distance

- ▶ For  $\mathbf{y} = (y_1, \dots, y_n) \in \{0, 1\}^n$ , let  $\|\mathbf{y}\|_1 = \sum_i y_i$  — Hamming weight of  $\mathbf{y}$
- ▶  $\|\mathbf{y} - \mathbf{y}'\|_1 = \|\mathbf{y} \oplus \mathbf{y}'\|_1$  — Hamming distance of  $\mathbf{y}$  from  $\mathbf{y}'$ ; # of places differ.
- ▶ We sometimes just write  $|\mathbf{y}|$ .



## Proving the theorem

- ▶ Fix  $p \in [0, \frac{1}{2})$  and  $\varepsilon > 0$ , and let  $m > m_\varepsilon$  and  $n \geq m(\frac{1}{C_p} + \varepsilon)$ , for  $m_\varepsilon$  to be determined by the analysis. (Recall  $C_p = 1 - h(p)$ ).
- ▶ We show  $\exists f: \{0, 1\}^m \mapsto \{0, 1\}^n$  and  $g: \{0, 1\}^n \mapsto \{0, 1\}^m$ , s.t.  
 $\Pr_{\mathbf{x} \leftarrow \{0, 1\}^m} [g(f(\mathbf{x}) \oplus \mathbf{Z}) \neq \mathbf{x}] \leq \varepsilon$
- ▶  $g(y)$  returns  $\operatorname{argmin}_{\mathbf{x}' \in \{0, 1\}^m} \|\mathbf{y} - f(\mathbf{x}')\|_1$
- ▶ So it all boils down to finding  $f$  s.t.  
 $\Pr_{\mathbf{x} \leftarrow \{0, 1\}^m; \mathbf{y} = f(\mathbf{x}) \oplus \mathbf{Z}} [\forall \mathbf{x}' \in \{0, 1\}^m \setminus \{\mathbf{x}\}: \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|_1 < \|\mathbf{f}(\mathbf{x}') - \mathbf{y}\|_1] \geq 1 - \varepsilon$
- ▶ Idea: for  $p' > p$  to be determined later, find  $f$  s.t. w.h.p. over  $\mathbf{x}$  and  $\mathbf{Z}$ :  
 (1)  $\|\mathbf{f}(\mathbf{x}) \oplus \mathbf{Z}, \mathbf{f}(\mathbf{x})\|_1 \leq p'n$   
 (2)  $\|\mathbf{f}(\mathbf{x}) \oplus \mathbf{Z}, \mathbf{f}(\mathbf{x}')\|_1 > p'n$  for all  $\mathbf{x}' \neq \mathbf{x}$



- ▶ We choose  $f$  uniformly at random (what does it mean?)
- ▶ Non-constructive proof
- ▶ Probabilistic method

## Proving there exists good $f$

- Fix  $p' > p$  such that  $0 < \frac{1}{c_{p'}} - \frac{1}{c_p} \leq \frac{\varepsilon}{2}$  and  $\mathbf{x} \in \{0, 1\}^m$
- For  $y \in \{0, 1\}^n$ , let  $B_{p'}(y) = \{y' \in \{0, 1\}^n : \|y' - y\|_1 \leq p'n\}$
- (1) By weak law of large numbers,  $\exists n' \in \mathbb{N}$  s.t.  $\forall n \geq n'$   
 $\alpha_n := \Pr_{z \leftarrow Z} [(f(\mathbf{x}) \oplus z) \notin B_{p'}(f(\mathbf{x}))] \leq \frac{\varepsilon}{2}$  (for any fixed  $f$ )
- Fact (proved later):  $b(p') := |B_{p'}(y)| \leq 2^{n \cdot h(p')}$

$$\Rightarrow \forall \mathbf{x}' \neq \mathbf{x} \in \{0, 1\}^m: \Pr_f[f(\mathbf{x}) \oplus Z \in B_{p'}(f(\mathbf{x}'))] = \frac{b(p')}{2^n} \leq \frac{2^{n \cdot h(p')}}{2^n} = 2^{-nC_{p'}}$$

$$\Rightarrow \beta_{m,n} := \Pr_f[\exists \mathbf{x}' \neq \mathbf{x} \in \{0, 1\}^m: f(\mathbf{x}) \oplus Z \in B_{p'}(f(\mathbf{x}'))] \leq 2^{m-nC_{p'}}$$

$$\Rightarrow \beta_{m,n} \leq \frac{\varepsilon}{2}, \text{ for } n \geq \frac{1}{c_{p'}}(m - \log \frac{\varepsilon}{2}) = m(\frac{1}{c_{p'}} - \frac{\log \frac{\varepsilon}{2}}{mC_{p'}})$$

$$(2) \beta_{m,n} \leq \frac{\varepsilon}{2}, \text{ for } m \geq m' := \left\lceil \frac{-\log \frac{\varepsilon}{2}}{\frac{\varepsilon}{2} \cdot C_{p'}} \right\rceil \text{ and}$$

$$n \geq m(\frac{1}{c_p} + \varepsilon) \geq m(\frac{1}{c_p} + \frac{\varepsilon}{2} + \frac{-\log \frac{\varepsilon}{2}}{mC_{p'}}) \geq m(\frac{1}{c_{p'}} - \frac{\log \frac{\varepsilon}{2}}{mC_{p'}})$$

- Fix  $m > m_\varepsilon := \max\{m', n'\}$  and  $n > m(\frac{1}{c_p} + \varepsilon)$ .

- $\Pr_f[g(f(\mathbf{x}) \oplus Z) \neq \mathbf{x}] \leq \alpha_n + \beta_{m,n} \leq \varepsilon$ .

$$\Rightarrow \exists f \text{ s.t. } \Pr_{\mathbf{x} \leftarrow \{0,1\}^m} [g(f(\mathbf{x}) \oplus Z) \neq \mathbf{x}] \leq \varepsilon. \quad \square$$

## Tightness

Let  $X \leftarrow \{0, 1\}^m$ ,  $Z = (Z_1, \dots, Z_n)$ , for  $Z_1, \dots, Z_n$  iid  $\sim (1-p, p)$ , let  $f: \{0, 1\}^m \mapsto \{0, 1\}^n$ ,  $g: \{0, 1\}^n \mapsto \{0, 1\}^m$ , and let  $Y = f(X)$ .

### Theorem 2

Assume  $\Pr[g(Y) = X] \geq 1 - \varepsilon$ , then  $nC_p \geq m(1 - \varepsilon) - 1$ .

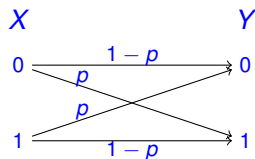
- ▶ Recall that **Thm 1** allows  $nC_p = m(1 + \varepsilon C_p)$ .
- ▶ Hence,  $\lim_{\varepsilon \rightarrow 0} \frac{m}{n} = C_p$

Proof:

- ▶  $\underbrace{X}_{m \text{ bits}} \rightarrow \underbrace{f(X)}_{n \text{ bits}} \rightarrow \underbrace{f(X) \oplus Z}_Y \rightarrow \underbrace{g(f(X) \oplus Z)}_{g(Y)}$
- ▶ By Fano,  $H(X|Y) \leq h(\varepsilon) + \varepsilon \cdot \log 2^m \leq 1 + \varepsilon m$
- ▶  $I(X; Y) = H(X) - H(X|Y) \geq m - \varepsilon m - 1 = m(1 - \varepsilon) - 1$
- ▶  $H(Y|X) = H(X, Y) - H(X) = H(X, Z) - H(X) = H(Z) = nh(p)$
- ▶  $I(X; Y) = H(Y) - H(Y|X) \leq n - nh(p)$
- ▶ Hence,  $m(1 - \varepsilon) \leq I(X; Y) + 1 = n(1 - h(p)) + 1 = nC_p + 1$

## Why $C_p = 1 - h(p)$ ?, intuitive explanation

- ▶ Let  $X \leftarrow \{0, 1\}$ ,  $Z \sim (1 - p, p)$  and  $Y = X \oplus Z$



- ▶  $I(X; Y) = H(Y) - H(Y|X) = H(Y) - H(Z) = 1 - h(p) = C_p$
- ▶ Received bit “gives”  $C_p$  information about transmitted bit
- ▶ Hence, to recover  $m$  bits, we need to send at least  $m \cdot \frac{1}{C_p}$  bits

A different proof:

- ▶ Let  $X \leftarrow \{0, 1\}^m$ ,  $Y = f(X)$  and assume that  $g(Y) = X$ .
- ▶ Hence,  $n \geq H(Y) = H(X, Z) = m + nh(p)$

$$\implies n(1 - h(p)) = nC_p \geq m$$

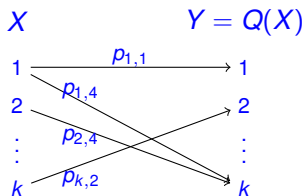
## General communication channel

$Q: [k] \mapsto [k]$  that channel (a probabilistic function)

$$p_{i,j} = \Pr[Q(i) = j]$$

- ▶  $\mathbf{x} = (x_1, \dots, x_m) \in \{0, 1\}^m$
- ▶ Encoding function  $f: \{0, 1\}^m \mapsto \{1, \dots, k\}^n$
- ▶ Decoding function  $g: \{1, \dots, k\}^n \mapsto \{0, 1\}^m$
- ▶  $\mathbf{x} \xrightarrow{\text{encoding}} f(\mathbf{x}) \xrightarrow{\text{channel}} Q(f(\mathbf{x})) \xrightarrow{\text{decoding}} g(Q(f(\mathbf{x})))$
- ▶ We hope for  $g(Q(f(\mathbf{x}))) = \mathbf{x}$

- ▶ **Channel capacity**  $C_Q = \max_p C(p)$ , for  $c(p) := I(X_p; Q(X_p))$  and  $X_p \sim p$ .
- ▶ The maximal information  $Y$  gives on  $X$
- ▶ Shannon theorem:  $\forall Q$  and  $\forall \varepsilon > 0$ ,  $\exists m_\varepsilon$ :  $\forall m > m_\varepsilon$  and  $\forall n > m(\frac{1}{C_Q} + \varepsilon)$ :  $\exists f, g$  as above s.t.  $\Pr_Q[g(Q(f(\mathbf{x}))) \neq \mathbf{x}] \leq \varepsilon$ , for all  $\mathbf{x} \in \{0, 1\}^m$ .
- ▶ Proof: similar lines to the binary case, but more subtle distribution for  $f$



## Discussion

- ▶ Tight result
- ▶ Non-constructive
- ▶ Coding theory: design **explicit** (and efficient) code achieving the same bounds
- ▶ Application: faulty communication, storage
- ▶ Combination of data compression and ECC

# Part II

## Hamming Ball

## Why $H(X_1, \dots, X_n) \leq \sum_i H(X_i)$ so useful?

- ▶  $\mathcal{S} \subseteq \{0, 1\}^n$ ;  $X = (X_1, \dots, X_n) \leftarrow \mathcal{S}$
- ▶  $\log |\mathcal{S}| = H(X) \leq \sum_i H(X_i)$  implies  $|\mathcal{S}| \leq 2^{\sum_i H(X_i)}$
- ▶ If  $\sum_i H(X_i)$  is small, then  $\mathcal{S}$  is small  
 $X_i$  are unbalanced, e.g.,  $\sim (0.1, 0.9)$ , implies  $|\mathcal{S}| \leq 2^{n \cdot h(0.1)} \leq 2^{n/2}$
- ▶  $\mathcal{S}$  is large implies  $\sum_i H(X_i)$  is large, hence most  $X_i$  are almost balanced
- ▶  $|\mathcal{S}| \geq 2^n/2$  implies  $\mathbb{E}_{i \leftarrow [n]} [H(X_i)] \geq 1 - \frac{1}{n}$
- ▶ Most  $X_i$  are close to uniform



# Hamming ball

- ▶  $p \leq \frac{1}{2}$ ;  $\mathcal{S} = \{(a_1, \dots, a_n) \in \{0, 1\}^n : \sum_i a_i \leq pn\}$
- ▶  $|\mathcal{S}| = \sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k}$
- ▶  $X = (X_1, \dots, X_n) \leftarrow \mathcal{S}$
- ▶  $\sum_i X_i \leq pn \implies \mathbb{E}[\sum X_i] \leq pn$ , and by symmetry  $\mathbb{E}[X_i] \leq p$  for every  $i$
- ▶ Hence,  $\Pr[X_i = 1] \leq p$  for every  $i$ .

$\implies H(X_i) \leq h(p)$  for every  $i$

$\implies |\mathcal{S}| \leq 2^{\sum_i H(X_i)} \leq 2^{nh(p)}$

$\implies \sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k} \leq 2^{nh(p)}$

## Corollary 3

For  $y \in \{0, 1\}^n$  and  $p \in [0, \frac{1}{2}]$ , let  $B_p(y) = \{y' \in \{0, 1\}^n : \|y' - y\|_1 \leq pn\}$ .  
Then  $|B_p(y)| = \sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k} \leq 2^{n \cdot h(p)}$

Very useful estimation. Weaker variants follows by AEP or Stirling,

## Hamming ball, cont.

The above bound yields the following concentration bound:

### Corollary 4

Let  $X_1, \dots, X_n$  be iid uniform bits and let  $p \in [0, \frac{1}{2}]$ , then  
$$\Pr [\sum_i X_i \leq pn] = \Pr [(X_1, \dots, X_n) \in \mathcal{S}] \leq 2^{nh(p)} \cdot 2^{-n} = 2^{-n(1-h(p))}.$$

Very useful inequality. No Chernoff just IT

# Part III

## **Combinatorial Applications**

# Movies

- ▶  $2^n$  people,  $m = 3n$  movies.
- ▶ Every pair of movies was seen by at least 90% of the people
- ▶ Claim: there exist two people who saw exactly the same set of movies
- ▶  $X \leftarrow [2^n]$  — a random person
- ▶  $g_i(x) = \begin{cases} 1, & x \text{ saw movie } i; \\ 0, & \text{otherwise.} \end{cases}$
- ▶  $Y_i = g_i(X)$
- ▶  $\forall i, j: H(Y_i, Y_j) \leq H(0.9, \frac{0.1}{3}, \frac{0.1}{3}, \frac{0.1}{3}) \leq \frac{2}{3}$
- ▶  $H(Y = (Y_1, \dots, Y_m)) \leq H(Y_1, Y_2) + H(Y_3, Y_4) + \dots + H(Y_{m-1}, Y_m) < \frac{3n}{2} \cdot \frac{2}{3} = n = H(X)$
- ▶ Hence,  $X$  is not determined by  $Y$

# Isoperimetric inequality

- ▶  $\mathcal{S} \subseteq \{0, 1\}^n$
- ▶ Edges of  $\mathcal{S}$  —  $E = \{(u, v) \in \mathcal{S} : |u - v| = 1\}$

## Theorem 5

$$|E| \leq \frac{1}{2} \cdot |\mathcal{S}| \cdot \log |\mathcal{S}|$$

- ▶ Equality if  $\mathcal{S}$  is “face” :  $\mathcal{S} = \{(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in \{0, 1\}^d\}$  for some  $\mathbf{x} \in \{0, 1\}^{n-d}$
- ▶ Example:  $\mathcal{S}$  is a **face** of the 3-dimensional cube  
 $n = 3, |\mathcal{S}| = 4$ , implies  $|E| \leq \frac{1}{2} \cdot 4 \cdot \log 4 = 4$
- ▶  $E_i$  — edges of  $E$  in **direction**  $i$  ( $E = \bigcup_{i \in [n]} E_i$ )
- ▶  $X = (X_1, \dots, X_n) \leftarrow \mathcal{S}$  and  $X_{-i} = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$

## Lemma 6

$$H(X_i | X_{-i}) = \frac{2|E_i|}{|\mathcal{S}|}$$

Proving **Thm 5**:

$$\begin{aligned} \log |\mathcal{S}| = H(X_1, \dots, X_n) &= H(X_1) + H(X_2 | X_1) + \dots + H(X_n | X_1, X_2, \dots, X_{n-1}) \\ &\geq H(X_1 | X_{-1}) + H(X_2 | X_{-2}) + \dots + H(X_n | X_{-n}) = \sum \frac{2|E_i|}{|\mathcal{S}|} = \frac{2|E|}{|\mathcal{S}|}. \quad \square \end{aligned}$$

## Proving Lemma 6

We prove for  $i = 1$

- ▶  $\mathcal{S} \subseteq \{0, 1\}^n$ ;  $X = (X_1, \dots, X_n) \leftarrow \mathcal{S}$
- ▶  $E = \{(u, v) \in \mathcal{S} : |u - v| = 1\}$  and  $E_1$  contains edges of  $E$  in direction 1
- ▶  $\mathcal{S}_{-1} := \{\mathbf{y} \in \{0, 1\}^{n-1} : \exists x \in \{0, 1\} \text{ s.t. } (x, \mathbf{y}) \in \mathcal{S}\}$ .  
( $\mathcal{S}$  projected on  $(2, \dots, n)$ )
- ▶  $\mathcal{S}_{-1}^e := \{\mathbf{y} \in \{0, 1\}^{n-1} : (0, \mathbf{y}), (1, \mathbf{y}) \in \mathcal{S}\}$  and  $\mathcal{S}_{-1}^{\neg e} = \mathcal{S}_{-1} \setminus \mathcal{S}_{-1}^e$
- ▶  $|\mathcal{S}| = 2|\mathcal{S}_{-1}^e| + |\mathcal{S}_{-1}^{\neg e}|$
- ▶  $|E_1| = |\mathcal{S}_{-1}^e|$
- ▶  $H(X_1|X_{-1}) = \Pr[X_{-1} \in \mathcal{S}_{-1}^e] \cdot 1 = \frac{2|\mathcal{S}_{-1}^e|}{2|\mathcal{S}_{-1}^e| + |\mathcal{S}_{-1}^{\neg e}|} = \frac{2|E_1|}{|\mathcal{S}|}$  □
- ▶ ...